California Department of Education

Measures for a College and Career Indicator: Research Brief on SAT & ACT

April 4, 2014

Educational Policy Improvement Center (EPIC)

Introduction

In September of 2012, Governor Jerry Brown signed into law Senate Bill 1458, which calls for California's school accountability system to shift from a near exclusive reliance on state test scores to a broader range of measures demonstrating student achievement. At the high school level, starting in the 2015–2016 school year, the Academic Performance Index (API) will include an indicator composed of measures reflecting students' college and career preparedness.

To determine exactly what measures will be included in this new indicator, the State Superintendent of Public Instruction and the State Board of Education will consider input from regional public meetings, a statewide survey, and recommendations from the Public Schools Accountability Act (PSAA) Advisory Committee. To further support this decision-making process, the California Department of Education has contracted with the Educational Policy Improvement Center (EPIC) to conduct analyses of six different types or clusters of potential measures of college and career preparedness, summarized in a series of six white papers and a final summary report.

This white paper considers college admissions exams—specifically the SAT® and ACT®—as potential measures to be included in California's College and Career Indicator (CCI). The paper begins by presenting a brief overview of the two exams, their respective histories, and their current applications to other state accountability systems. Next, the SAT and ACT are evaluated against an analytical framework to determine the technical quality, stakeholder relevance, and system utility of tests when used as a component measure of accountability. The paper concludes with a summary of the analysis, identifying major strengths, weaknesses, and tradeoffs.

The SAT and ACT: A Brief Overview

The SAT was initially known as the Scholastic Aptitude Test. It was first administered as an experimental multiple-choice exam in 1926. Born out of the IQ testing movement, it was based on the premise that a complex set of human behaviors could be reduced to a series of discrete elements, that each of those elements could then be measured independently, that performance on all items could then be summed to generate an approximation of an individual's capabilities on the complex behaviors in question, in this case mental capacities associated with college. The exam was designed to be easily scored and to assess students' general analytic ability rather than content knowledge. When Harvard University adopted the exam in 1933, its primary goal in doing so was to use the exam as a tool for equity—to identify promising applicants from outside the privileged pipeline of private, largely east coast, boarding schools. By the end of World War II, the SAT had displaced the essay form of admissions testing that the College Board had been administering in parallel with the SAT. This happened in part because of the difficulty of scoring essay exams during World War II, as the scorers were not always able to drive to the scoring sites due to gas rationing. Therefore, the multiple-choice form was rapidly adopted by all member institutions of the College Board as a standardized measure of student ability (Lemann, 1999).

Over time, the College Board has moved way from "aptitude" in both name and intention. The test became the Scholastic Achievement Test for a period of time in the late 1980s and early 1990s, before the College Board simply trademarked the term SAT with no underlying meaning for the letters as an acronym. The instrument continued to focus on "developed reasoning" skills that may

or may not occur as a result of what is taught in the classroom. Since 2005, the SAT has consisted of three sections, each worth up to 800 scaled points, resulting in a maximum score of 2400 points. The sections are mathematics (SATM), verbal (SATV), and writing (SATW). Students have 3 hours and 45 minutes to complete the entire test. In addition to administering the SAT, the College Board also offers two other assessments at eighth and tenth grades, Readistep and PSAT, and up to 20 SAT subject tests at an additional cost. However, as noted below, the SAT is about to undergo a major redesign in 2016.

The ACT was developed in 1959. Rather than assessing general analytic ability, this college admissions exam was designed to focus on content taught in classrooms and student achievement relative to that content. Its founder, E. F. Lindquist, explained the proper signaling role of the exam for both students and colleges: "If the examination is to have the maximum motivating value for the high school student, it must impress upon him the fact that his chances of being admitted to college … depend not only on his 'brightness' or 'intelligence' or other innate qualities or factors for which he is not personally responsible, but even more upon how hard he has worked at the task of getting ready for college" (Atkinson & Geiser, 2009, p. 7). As such, the ACT assesses content knowledge and skills gleaned from surveys of high school and college instructors nationally and, more recently, the ACT College Readiness Standards.

The ACT comprises four sections: English (ACTE), mathematics (ACTM), reading (ACTR), and science (ACTS), and an optional writing section (ACTW). Maximum score in each section is 36 points. ACT scores are not commonly summed across subject-area tests, as are SAT scores. Students are given 2 hours and 55 minutes to complete the English, mathematics, reading, and science portions, and have 30 minutes to complete the optional ACTW. ACT administers the ACT test as the terminal exam of its college readiness suite that consists of EXPLORE at grade 8 and PLAN at grade 10. The company also offers a range of other tests and products, most notably WorkKeys, designed to assess career readiness.

The SAT and ACT are now taken by approximately the same number of students throughout the United States (Atkinson & Geiser, 2009). Although the exams historically were more commonly taken and accepted in particular geographic regions, both are now accepted and used by nearly every four-year institution in the country. The two tests now have more similarities than differences. The SAT has undergone several revisions to situate its reasoning tasks within the course content a student might encounter in high school, and many ACT tasks and test items now require less content knowledge and more analytic ability. In March 2014, the College Board announced a set of planned revisions to be implemented in the 2016 administration of the SAT. These include making the SATW optional, removing penalties for wrong answers, replacing obscure words in the vocabulary with words students commonly encounter in college courses, making reading passages longer and more closely derived from what students read in college courses, including more informational texts in addition to literature, having students react to excerpts from "foundational documents" of Western democracy such as the Constitution and the Gettysburg Address, covering fewer topics in depth on the SATM, and removing sentence completion questions (College Board, 2014b). While each exam retains its own philosophical perspective on college preparedness, the basic underlying premise of each admissions exam is to measure student preparedness for college and to do so with measures that are actionable by students, that students can learn how to do and are not dependent entirely on natural ability or privileged opportunity.

With an increased national policy focus on college and career preparedness, both ACT and the College Board have developed benchmark scores that policymakers and administrators can use to evaluate the postsecondary preparedness of groups of students. Using a sample of approximately 68,000 students across 110 four-year institutions, the SAT college readiness benchmark of 1550 out of 2400 "indicates a 65 percent likelihood of achieving a B- average or higher during the first year of college"[1]. Using a similar sample and correlation method, the ACT benchmarks represent the "level of achievement required for students to have a 50% chance of obtaining a B or higher or about a 75% change of obtaining a C or higher in corresponding credit-bearing first-year college courses." The college readiness benchmark scores are as follows: ACTE, 18; ACTM, 22; ACTR, 22; and ACTS, 23. It is important to note that both testing organizations warn that these benchmark scores are appropriate for evaluating the general readiness groups of students; benchmarks should not be used as cut score determinations of an individual student's readiness. However, these warnings have not stopped some postsecondary institutions from using these indicators as cut scores to make summative judgments about prospective students.

States are increasingly using or adapting these preparedness benchmarks as tools for accountability. As of March 2014, 25 states currently use or plan to incorporate the SAT, ACT, or both into their high school accountability systems. The role of the exams and the treatment of student scores, however, vary across state systems. Some states treat the ACT or SAT as a universal measure of preparedness, requiring all students to take college admissions exams regardless of their postsecondary aspirations. For example, Colorado requires all 11th graders to take the ACT and uses the state average ACT score as the benchmark level of readiness. Schools receive four points if the average ACT score for their students is at or above 22, three points if the average score is above the state average (20.4 in 2013) but below 22, two points if the average score is at or above 17 but below the state average, and one point if the average score is below 17. Average ACT scores account for 25% of a high school's College and Career Readiness Indicator alongside graduation rate (25%), disaggregated graduation rate (25%), and dropout rate (25%). The College and Career Readiness Indicator, in turn, accounts for 35 out of 100 points in a high school's overall quality rating, alongside achievement (15%), growth (35%), and growth gaps (15%) observed in the state assessments.

Other states treat SAT and ACT scores as conditional measures, representing one of several data points that may be included in a high school rating. For example, Kentucky implemented its Next Generation accountability system in 2013, evaluating schools across three general categories: Next Generation Learners (accounting for 70% of overall school rating), Next Generation Instruction and Support (20%), and Next Generation Teachers and Leaders (10%). At the high school level, the Next Generation Learners category includes a college and career readiness rate, alongside achievement, growth, gap, and graduation rates, each representing a 20% category score. The readiness rate is a percentage calculated by dividing the number of students who have successfully met a CCR indicator by the total number of students who graduated. Students may indicate college readiness by meeting or exceeding state benchmarks on the ACT, COMPASS, or state placement exam. Students may indicate career readiness by passing state benchmarks on a general knowledge and skills exam (ACT WorkKeys or Armed Services Vocational Aptitude Battery), as well as demonstrating specific technical knowledge and skills through industry certification or passing a state CTE exam. Students may also indicate college and career readiness, adding a half-point bonus

[1] http://pathway.collegeboard.org/data-and-reports

to the school's readiness rate. Kentucky has set readiness benchmarks for ACTR (20), ACTE (18), and ACTM (19).

Each of these accountability approaches carries significant implications for the quality, relevance, and utility of a college admissions exam as a measure of high school quality. The following section explores these issues more thoroughly, evaluating the SAT and ACT against a set of criteria by which state decision makers can consider their potential application to the California school accountability system.

Evaluation Against an Analytical Framework

Working in collaboration with the PSAA Advisory Committee, EPIC developed an analytical framework to provide a consistent, rigorous set of criteria by which each measure can be evaluated for its inclusion in the API. This framework was adapted from the Advisory Committee's API Guiding Principles and was supplemented with additional criteria specific to the charge of designing a College and Career Indicator (CCI). Organized under the dimensions of technical quality, stakeholder relevance, and system utility, the following 10 criteria explore the extent to which each measure under consideration

- has a research base demonstrating a relationship with postsecondary success;
- allows for fair comparisons;
- is stable;
- has currency outside the accountability system;
- is understandable to the public;
- measures content, skills, and competencies that can be taught and learned in school;
- emphasizes student performance, not educational processes;
- minimizes burden;
- includes as many students as possible; and
- recognizes a variety of postsecondary pathways.

The design of the framework acknowledges that satisfaction of the above criteria is not a simple binary decision of yes or no. Analyses will be nuanced, supported by research, and summarized on a consistent scale or choice set applied across all six clusters of measures considered in this white paper series. Additionally, analyses may sometimes place criteria in conflict with one another (e.g., a measure may have a strong evidence base but place an extraordinary implementation burden on schools). The purpose of this work is not to make recommendations, but rather to provide decision makers with the necessary information to identify the strengths, weaknesses, and trade-offs associated with each measure considered for inclusion in the College and Career Indicator.

The following subsections evaluate the SAT and ACT against the analytical framework, taken as both distinct and unique instruments and as a general cluster or class of college admissions exams.

A. Technical Quality

For the purposes of this research review, technical quality is defined as having predictive validity for forecasting how students will perform in postsecondary pathways, allowing fair comparisons among

different subpopulations of students, and having sufficient stability to allow for examination of trends.

A1. Relationship to Postsecondary Success

The first of the 10 evaluative criteria looks at the empirical research base to explore the relationship between the measure and postsecondary success. For the purposes of this project, research on postsecondary success may include a wide array of outcome variables including college matriculation, persistence, course grades, grade point average, and degree completion. Career success outcome variables may be defined extrinsically (e.g., salary or promotion) or intrinsically (e.g., self-reported job satisfaction). The evidence base for each measure or cluster of measures is evaluated on a four-point scale: *no evidence,* or *weak, moderate,* or *strong* relationships.

Both the College Board and ACT have conducted extensive research to establish the validity of their exams as predictors of postsecondary success (Camara & Echternacht, 2000; Kobrin et al., 2008; Morgan, 1989; Noble and Sawyer, 2002; Sanchez, 2013). These studies generally employ simple correlation methods and identify freshman year grade point average (FYGPA) as the postsecondary success outcome variable. These studies have generated consistent findings: (1) SAT/ACT scores and high school grade point average (HSGPA) are positively correlated with FYGPA, (2) when HSGPA is not included in the analyses, the predictive power of the SAT/ACT decreases, and (3) the SAT/ACT combined with HSGPA produces the highest correlation with postsecondary success. The fact that the exams' predictive power is strongest when combined with other student data suggests that the SAT and ACT accomplish what the exams are designed to do: they provide common measures of probability of success in college that can be considered along with a number of other student characteristics to inform college admissions decisions.

While outside researchers have replicated the findings of the College Board and ACT studies, others have been critical of the simple correlational methods employed by this body of research. In other words, the magnitude of the SAT and ACT's contribution in predicting college success has been found to vary based on the independent variables included or controlled for in the design of the study (Atkinson & Geiser, 2009; Berry & Sackett, 2009; Bettinger, Evans, & Pope, 2011; Geiser & Studley, 2002; Grissmer, 2000; Rothstein, 2004; Zwick, Brown, & Sklar, 2004). Numerous studies have found that student and school socioeconomic status (SES) and SAT/ACT scores are strongly correlated (Buchmann, Condron, & Roscigno, 2010; Byun & Park, 2012; Card & Rothstein, 2007; Rothstein, 2004; Zwick & Himelfarb, 2011; Zwick, 2012). Rothstein (2004) analyzed data from seven University of California institutions and found that omitting SES from validity models produced findings that overstated the SAT's role in predicting postsecondary success by as much as 150%. When SES is controlled for, Rothstein contends the incremental validity attributed to the SAT justifies the use of the exam's score in college admissions.

Multiple research studies have also shown that the SAT and ACT subject tests or even specific sections are better predictors than the traditional exams taken as a whole (Atkinson & Geiser, 2009; Geiser & Studley, 2002; Zwick, 2004). Bettinger, Evans, and Pope (2011) found that ACTR and ACTS, taken individually, have essentially no predictive power and that the exam's predictive power is found in the ACTM and ACTE sections. A number of studies have shown that the SATW was more effective and consistent than the SATM or SATV in predicting student performance (Kobrin et. al., 2008; Zwick, Brown, & Sklar, 2004). These findings take on even greater significance

considering the College Board's proposed changes to the 2016 exam making the writing component optional for test takers.

Despite these issues, researchers generally agree that SAT and ACT scores positively correlate to postsecondary success, but not to an overwhelming degree. They provide useful information that can and should be used in combination with multiple additional data sources. The evidence that these exams are useful measures of college preparedness is at the moderately strong level. Note, however, that this may not be as true for career preparedness, depending in part on how career preparedness is defined. The forthcoming changes to the SAT do raise some more global questions as to the revised exam's predictability. Correlation studies require at least two years of data (i.e., the student's senior year in high school to take the exam and freshman year in college to establish a grade point average). The very nature of the proposed redesign, however, is driven by an evidence-based model that derives from studies of what is actually taught in freshman college courses, so it is reasonable to assume that the revised exam will predict at least as well as the current version.

A2. Fair Comparisons

This evaluative criterion is based on the assumption that the API must give all students a fair chance to show what they know and have learned. For the purposes of this study, the extent to which a measure provides fair comparisons across students and schools is determined by careful attention to bias and summarized on a three-point scale: the measure fully *allows* for fair comparisons, *partially does*, or *does not*.

Recalling the fact that the SAT was originally conceived as a tool for equity and opportunity and the ACT as a measure of what was commonly taught in high school classrooms, both college admissions exams ironically have long been criticized for bias against certain subgroups of students. As stated in the previous section, research has shown that students' socioeconomic status (SES) is strongly correlated with SAT/ACT scores (Rothstein, 2004; Sackett, Kuncel, Arneson, Cooper, & Waters, 2009; Zwick, 2012). School-level composite SES has also been found to influence aggregated SAT/ACT scores (Zwick & Himelfarb, 2011). Student race adds another dimension of concern, where family poverty negatively affects the scores of African-American test takers more than their white peers after controlling for other factors (Dixon-Román, Everson, & McArdle, 2013). Reports from the testing companies themselves identify racial gaps in student scores. In a study recently released by ACT, only 5% of African American students met ACT readiness benchmarks on all four sections of the exam, as compared to 26% of the general test-taking population (ACT, 2013b).

The College Board and ACT often cite these kinds of statistics as reflections of the lack of comparable opportunities and access to rigorous coursework for students in all high schools nationally. The test makers see these differences in subgroup performance as indicators of system pathology rather than biases inherent in their assessment instruments. While it is true that one of the strongest predictors of postsecondary success is a student's SES, and while it is not uncommon to find double-digit achievement gaps in other educational outcome data, research has shown that the language and logic of test items themselves may also hide some culturally based bias (Freedle, 2003; Santelices & Wilson, 2010). This criticism is controversial and has been strongly contested by the College Board and ACT, as both organizations regularly issue evidence of rigorous review processes to check for and prevent such item-level bias (Camara & Sathy, 2004; College Board, 2010). However, the differences in access to quality instructional programs across high schools by racial, ethnic, and economic groups are very real (Geiser & Santelices, 2004). Differences in performance

across groups are likely an admixture of subtle cultural biases and opportunity factors rather than being simply one or the other.

Gender is a student characteristic that further complicates fair comparisons. Females tend to have higher HSGPAs than males but do not score as well on the SAT or ACT, a phenomenon known as the female underprediction effect (FUE). Admissions tests underpredict women's and overpredict men's academic performance (Kling, Noftle, & Robins, 2013). A possible reason for this underprediction is that high school grades are the result of many factors, such as studiousness, attitudes, academic preparation, conscientiousness and study skill, that go beyond content knowledge or analytic abilities (Kling, Noftle, & Robins, 2012; Stricker, Rock, & Burton, 1991).

Aggregating SAT and ACT scores within schools without the correct control variables and methodological adjustments make it problematic to compare student performance across schools. In state contexts where SAT and ACT are voluntary, comparisons across schools say more about the sample characteristics of self-selected test takers than about school quality (Grissmer, 2000). Students who take the SAT/ACT are likely to perform better in college than those who do not take these tests, producing higher observed mean scores than the mean would be if the exam was compulsory and scores were available for all students. For example, Colorado required all high school juniors take the ACT beginning in the spring of 2001. In 2001, prior to implementation, 62% of high school juniors took the ACT and the average composite score was 21.5. In 2002, post implementation, 99% took the ACT and the average score dropped to 20.1. Such a treatment of scores can even create the perverse incentive for schools to discourage some students from taking exams. On the other hand, when the exam is compulsory, making determinations based solely on ACT or SAT scores can underrepresent a school's overall preparedness by not recognizing achievements that prepare students for other postsecondary pathways (e.g., two-year technical degrees, apprenticeships, the military). Finally, documented correlations with SES, race, and income—whether a reflection of a broken education system or an indication of biased instruments—do suggest the need to control for demographics statistically when making school-level comparisons.

Given the interaction between the potential bias of the measures, their tendency to underrepresent the performance potential of some students, and their inability to account for differences in the educational quality of a student's program, this analysis finds that the SAT and ACT allow for partially fair comparisons of students and schools.

A3. Stability

This evaluative criterion is chiefly concerned with how the measure contributes to the comparability and flexibility of the API as a whole over time. In order to measure school performance and improvement consistently and comparably over time, all components of a measurement system should be based on definitions that remain relatively constant from year to year. Likewise, the core measures within the College and Career Indicator system need to be reasonably stable. If they are, then the API has some capacity to incorporate future component measures of preparedness, which is important due to the dynamic nature of college and career preparedness. The stability of each measure or cluster of measures is evaluated here on a three-point scale: *not stable, partially stable*, and *fully stable*.

The SAT and ACT ensure year-to-year comparability of test forms through ongoing equating studies. During periods of time following the redesign of either exam, ACT and College Board have

conducted more rigorous equating studies to ensure the same underlying constructs were assessed and effects on subgroups were not exacerbated by the changes. A 2005 redesign of the SAT included the addition of a writing section (SATW) worth the same number of points as the SATV and SATM. This made the SAT maximum composite score 2,400 points instead of 1,600. Research from the College Board suggests that the post-2005 SAT is comparable with previous versions of the test (Kobrin et. al., 2008). Outside researchers, however, have argued that the SATW did alter the predictive validity of the test, and thus overall comparability between 2004 and 2005 versions (Kobrin, et. al, 2008; Zwick, Brown, & Sklar, 2004). The next SAT revision in 2016 will be accompanied by new equating and comparability studies. ACT has not announced any plans for a major redesign of its exam. The last substantial revision of the ACT took place in 1989, when all four content sections were revised and two sections—science and reading—changed to their current names. The ACT added an optional writing section in 2005, mirroring changes the SAT made in the same year.

Both SAT and ACT exams remain stable from year to year, with occasional test revisions that are closely monitored and measured. The underlying constructs generally remain unchanged or consistent, and revisions have historically improved the exams' alignment with the needs, processes, and content of postsecondary education.

B. Stakeholder Relevance

Accountability measures that are relevant to a variety of education stakeholder groups for more purposes than solely rating a school or district provide greater value to the levels of the education system than measures that meet only school and district accountability requirements. To the extent measures can serve multiple purposes, they may help increase stakeholder acceptance of an accountability system.

B1. Student Currency

This evaluative criterion is chiefly concerned with the extent to which component measures of the College and Career Indicator (CCI) are likely to be actionable and accepted by students. Rather than an assessment or data point that is only valuable in making system-level determinations of school quality, a CCI that has student currency reflects and creates incentives for behaviors and performances that directly affect or improve an individual student's prospects for success after high school.

The SAT and ACT have strong currency for students who aspire to attend four-year institutions because they satisfy a common admissions requirement. Research from Hyman (2013) and Klasik (2011) shows that states that require students to take the SAT or ACT see increases in enrollment at four-year postsecondary institutions. The policy impact on enrollment is directly related to the fact that the exam does have currency with students. Those students who do not view themselves as "college-going material" are not likely to take a college admissions exam voluntarily. When required or given effective incentives to do so, however, their resultant scores may make them more aware or open to considering a four-year program among their postsecondary options.

It should be noted that a limited number of colleges and universities are making admissions tests optional for applicants. FairTest, an advocacy group with a mission of critiquing standardized testing, keeps a current list of institutions that are either "test optional," "test flexible," exempt students from submitting test scores if they have qualifying HSGPA, or use admissions test scores

for research purposes only. As of 2013, 815 four-year institutions, 46 of which are in California, have made admissions tests optional, according to FairTest (2014). These 815 institutions represent 25% of the 3,217 four-year postsecondary institutions in the United States (National Center for Education Statistics, 2014). It is worth noting, however, that many are small liberal arts schools that have always examined a wide range of information when making admissions decisions, and many others are open enrollment or low-selectivity institutions. While a few high-visibility selective institutions, once again mostly smaller liberal arts institutions, have gained notoriety for becoming admissions test-optional, few larger or highly selective institutions have adopted this policy.

B2. Public Understanding

The API is intended to give educational stakeholders—educators, parents, students, and the public at large—a clear picture of a school's status and growth. The College and Career Indicator should therefore clearly communicate how it supports college and career preparedness in a way that is easily understood by non-educators as well as educators.

The College Board and ACT provide web pages that help students, teachers, educators, and policymakers understand SAT/ACT scores. These webpages include descriptions of how scores are calculated, sample score reports, details about how to interpret percentiles, and other resources (College Board, 2014a, 2014c; ACT, 2014b). The percentile rank, which indicates the percentage of students who scored below a given score point, is easier for most people to understand than raw scores. This information allows students, educators, and policymakers to make comparisons across schools, states, and the nation. Likewise, most colleges publish average SAT and ACT scores and the range of those scores for enrolled freshman. This helps prospective students to gauge whether a given college or university is a "good fit" for them.

As described previously, both the College Board and ACT have benchmark scores that policymakers and administrators can use to evaluate college preparedness for groups of students. Using a sample of approximately 68,000 students across 110 four-year institutions, the SAT college readiness benchmark of 1550 out of 2400 "indicates a 65 percent likelihood of achieving a B- average or higher during the first year of college." Using a similar sample and correlation method, the ACT benchmarks represent the "level of achievement required for students to have a 50% chance of obtaining a B or higher or about a 75% change of obtaining a C or higher in corresponding credit-bearing first-year college courses" (ACT, 2013b). The college readiness benchmark scores are as follows: ACTE, 18; ACTM, 22; ACTR, 22; and ACTS, 23. Again, it is important to note that both testing organizations warn that these benchmark scores are appropriate for evaluating the general preparedness of groups of students; benchmarks should not be used as cut score determinations of an individual student's preparedness

This analysis finds that the SAT and ACT admissions exams meet the criterion that the measure be understandable to educators, parents, students, and the public at large.

B3. Content, Skills, and Competencies

In order for the API to provide a valid description of school quality, its component parts must measure content, skills, and competencies that are taught and learned in schools. This criterion—evaluated on a three-point scale—addresses not just the validity of the accountability measure but also the actionability of a College and Career Indicator.

The SAT and ACT are both rooted in academic subject areas yet address student performance and potential in different ways. The SAT has historically been considered a developed reasoning test that measures the kinds of general analytic abilities required by college coursework that can be acquired in and outside of classroom through rigorous coursework and independent individual actions, such as recreational reading. The ACT has relied on national curriculum surveys to emphasize student achievement of content knowledge and skills.

The advent of the Common Core State Standards (CCSS), now adopted by 44 states and the District of Columbia, does enhance the opportunity for alignment between high school and college. This opens the door for increased alignment of admissions exams with high school coursework. The standards themselves were developed with involvement of both K–12 and postsecondary content experts who shared expectations of the knowledge and skills students need to succeed after high school. An independently conducted validity study found the standards to be overwhelmingly important and applicable to college courses, as rated by entry-level college faculty (Conley et al., 2011). An alignment study conducted by ACT found that the ACT is strongly aligned with the CCSS, although no independent research to date has validated this study (ACT, 2010). The College Board's plans for the redesign of the SAT largely echo the features of the CCSS: attention to text complexity, the use of informational texts from a variety of disciplines, the ability to use evidence to support arguments, the ability to use mathematical concepts in a variety of disciplinary contexts to solve problems, and others. Validating the degree of alignment with the CCSS can begin once the redesigned test is released in 2016.

Beyond potential alignment with the Common Core State Standards, the school-level actionability of the SAT and ACT is enhanced by the potential to use complementary assessments at the 8th and 10th grades. These link to their respective college admissions exam, allowing for score predictions and score reports that students, parents, and educators can use to plan actions to improve college preparedness.

Grounded in course content, potentially aligned to CCSS, and embedded in a suite of actionable assessments, the SAT and ACT are reasonable measures of key knowledge, skills, and competencies taught and learned, particularly in English and mathematics.

B4. Emphasis on Student Performance

The legislative charge to California's school accountability system is to focus on educational outcomes rather than inputs. As important as it is to account for different features of quality schooling (e.g., teachers, instructional resources, curriculum, and school organization), this evaluative criterion looks at the extent to which potential component measures of the College and Career Indicator emphasize student performance.

As college admissions exams are administered to individual students, both the SAT and ACT directly measure student performance and not educational inputs or processes.

C. System Utility

Measures to be included in an accountability system have greater utility if they add minimal burden to the education system yet include as many students as possible. The measures also are most useful when they are applicable to students who will pursue a variety of postsecondary pathways.

C1. Minimal Burden

Minimizing the burden of component measures of the College and Career Indicator means constraining the time and cost of implementation and data collection processes to the maximum extent possible. This criterion considers direct and indirect effects, for example, time to take a test and instructional time devoted to test prep, and the effects on students, teachers, administrators, and the system as a whole.

As voluntary college admissions exams, it costs a student $51 to take the SAT and $52.50 to take the ACT, including the optional writing exam ($36.50 without). The fees for both exams include sending up to four score reports to colleges. Students pay $11.25 for each additional SAT score report or $12 for each additional ACT score report sent to colleges. Also, both the College Board and ACT provide fee waivers to low-income students, and part of the 2016 SAT redesign plan includes increased supports to low-income students in the form of free score reports. Direct time burdens for the students are under four hours. The indirect burden of test preparation, however, is a potential factor. The billion-dollar test-prep industry introduces an additional variable when considering the relationship between family income and SAT or ACT score (Buchmann, Condron, & Roscigno, 2010; Byun & Park, 2012). The 2016 SAT redesign plans include a partnership with Khan Academy to provide free online test prep materials, but no measure is likely to eliminate entirely the market for students with the means and desire to try to gain advantage on the tests.

If incorporated into the state accountability system, the system-level requirements will include ordering scores from the College Board and ACT as well as aggregating and analyzing scores. The exact cost depends on how the scores are incorporated into the API, which determines the type of analysis needed. Research from Hyman (2013) suggests that the state cost in analyzing SAT or ACT scores is approximately $2 per student, but the author emphasizes that this cost varies from state to state. As a conditional rather than universal measure, burden would continue to be minimal for teachers, administrators, schools, and districts.

Considering these time and cost requirements along a continuum that includes other options, this analysis finds that the measure minimizes burden for students, educators, and the system as whole.

C2. Student Coverage

The API Guiding Principles state that the API should include as many students as possible in each school and district. This inclusion principle was cornerstone to an accountability system based entirely off universal measures (e.g., all students must take state assessments including populations requiring testing accommodations). The proposed College and Career Indicator is by necessity composed of conditional measures because not all students can be compelled to go to college, nor would it be desirable to do so. Students and their parents retain the right to choose which path makes the most sense for them, and college is only one option among many. In addition, students can demonstrate preparedness through an array of measures that are empirically linked to postsecondary success but that address different knowledge, skills, and aspirations. This evaluative criterion gives preference to scaled or scalable measures over local and unique ones.

The SAT and ACT are ubiquitous in California high schools, though not every high school student opts to take the test. In 2013, 234,767 Californians took the SAT and 107,243 took the ACT. The vast majority (88%) of California SAT test takers attended public high schools. The ACT data includes sophomores, juniors, and seniors, but does not disaggregate data by grade level nor whether

students attended public or private schools. When looking at California subgroup participation, white and Asian students are overrepresented and Hispanic students are underrepresented among both SAT and ACT test takers (California Department of Education, 2013; College Board, 2013; ACT, 2013a). These participation rates also closely mirror racial representation in four-year postsecondary institutions.

C3. Postsecondary Pathways

The last criterion is less an evaluation of a measure than a categorization to inform more global decisions about the API. A College and Career Indicator must include component measures that collectively or individually recognize a diverse set of postsecondary pathways. Thus, this criterion identifies whether a component measure supports a college-going pathway, career-going pathway, both, or neither.

The SAT and ACT support the college-going postsecondary pathway, and within that pathway the emphasis is on four-year institutions. The College Board states, "SAT and SAT Subject Tests are a suite of tools designed to assess your academic readiness for college" (College Board, 2014a). The ACT states that its test is a "curriculum- and standards-based educational and career planning tool that assesses students' academic readiness for college" (ACT, 2014a).

Summary Analysis

The SAT and ACT college admissions tests predict postsecondary success with a few caveats. Researchers and policymakers should be sensitive to the fact that school and demographic factors. When significant variables including SES, race, family income, gender, and sample size are controlled for in study designs, the relationship between the SAT/ACT and postsecondary success is not as strong across schools and subgroups of students. This means that comparisons need to be made with greater caution, and inclusion of the tests in an accountability system should be structured to allow for fair comparisons among schools, given differential effects of student opportunity factors and the tests' content sections on the relationship of the tests to postsecondary success.

The stability of the SAT and ACT is one of the strong points of this measure. The redesign undertaken by the College Board in 2005 was motivated in large measure by critiques of the test by then-president of the UC System Richard Atkinson, which demonstrates a responsiveness to California policy concerns. The current redesign demonstrates responsiveness to potential changes in teaching and learning heralded by the Common Core State Standards. This type of adaptation over time, closely monitored and researched, helps these tests be more valid and more stable than many other measures.

The SAT/ACT provides currency to students planning to attend a four-year postsecondary institution. For these students, the test contains currency because it is used to meet a college admissions requirement, though not all postsecondary institutions require an admissions test score. The score is most relevant for students planning to attend four-year universities.

Finally, SAT/ACT scores are reasonably understandable by students and parents, although school or district reports may be more potentially problematic and challenging to interpret properly. The cost of this measure mostly falls on the student, who is responsible for preparing for, taking, and

managing the score reports generated by the test. The CDE will incur some cost from ordering, aggregating, analyzing, and generating the data used to hold high schools accountable.

References

ACT. (2010). *The alignment of Common Core and ACT's college and career readiness system.* Retrieved from https://www.act.org/commoncore/pdf/CommonCoreAlignment.pdf

ACT. (2013a). *ACT profile report–state: Graduating class of 2013, California.* Retrieved from http://www.act.org/newsroom/data/2013/pdf/profile/California.pdf

ACT. (2013b). *The condition of college & career readiness 2013: African American students.* Retrieved from http://www.act.org/newsroom/data/2013/states/pdf/AfricanAmerican.pdf

ACT. (2014a). *The ACT test.* Retrieved from http://www.act.org/products/k-12-act-test

ACT. (2014b). *Understand your scores.* Retrieved from http://www.actstudent.org/scores/understand/#multchoice

Atkinson, R. C., & Geiser, S. (2009). Reflections on a century of college admissions tests. *Educational Researcher, 38*(9), 665–676.

Berry, C. M., & Sackett, P. R. (2009). Individual differences in course choice result in underestimation of the validity of college admissions systems. *Psychological Science, 20*(7), 822–830.

Bettinger, E. P., Evans, B. J., & Pope, D. G. (2011). *Improving college performance and retention the easy way: Unpacking the ACT exam* (NBER Working Paper No. 17119). Cambridge, MA: National Bureau of Economic Research.

Buchmann, C., Condron, D. J., & Roscigno, V. J. (2010). Shadow education, American style: Test preparation, the SAT and college enrollment. *Social Forces, 89*(2), 435–461.

Byun, S. Y., & Park, H. (2012). The academic success of East Asian American youth: The role of shadow education. *Sociology of Education, 85*(1), 219–239.

California Department of Education. (2013). *Fingertip facts on education in California: CalEdFacts.* Retrieved from http://www.cde.ca.gov/ds/sd/cb/ceffingertipfacts.asp

Camara, W. J., & Echternacht, G. (2000). *The SAT [R] I and high school grades: Utility in predicting success in college* (Research Notes RN-10). New York, NY: The College Board.

Camara, W. J., & Sathy, V. (2004). *College Board response to Harvard Educational Review article by Freedle.* New York, NY: The College Board.

Card, D., & Rothstein, J. (2007). Racial segregation and the black–white test score gap. *Journal of Public Economics 91*(11), 2158–2184.

College Board. (2010). *College Board response to Harvard Educational Review article by Santelices and Wilson.* Retrieved from http://research.collegeboard.org/sites/default/files/publications/2012/7/misc2010-1-response-harvard-ed-review-santelices-wilson.pdf

College Board. (2013). *State Profile Report: California.* Retrieved from http://media.collegeboard.com/digitalServices/pdf/research/2013/CA_13_03_03_01.pdf

College Board. (2014a). *About the tests: SAT and SAT subject tests overview.* Retrieved from https://sat.collegeboard.org/about-tests

College Board. (2014b). *The College Board announces bold plans to expand access to opportunity; Redesign of the SAT*. Retrieved from https://www.collegeboard.org/releases/2014/expand-opportunity-redesign-sat

College Board. (2014c). *Understanding your scores.* Retrieved from http://sat.collegeboard.org/scores/understanding-sat-scores

Conley, D. T., Drummond, K. V., de Gonzalez, A., Seburn, M., Stout, O., & Rooseboom, J. (2011). *Lining up: The relationship between the Common Core State Standards and five sets of comparison standards.* Eugene, OR: Educational Policy Improvement Center.

Dixon-Román, E. J., Everson, H. T., & McArdle, J. J. (2013). Race, poverty and SAT scores: Modeling the influences of family income on black and white high school students' SAT performance. *Teachers College Record, 115*(4).

FairTest. (2014). *Colleges and universities that do not use SAT/ACT scores for admitting substantial numbers of students into bachelor degree programs.* Retrieved from http://fairtest.org/sites/default/files/OptionalPDFHardCopy.pdf

Fleming, J. (2002). Who will succeed in college? When the SAT predicts black students' performance. *The Review of Higher Education, 25*(3), 281–296.

Freedle, R. O. (2003). Correcting the SAT's ethnic and social-class bias: A method for reestimating SAT scores. *Harvard Educational Review, 73*(1), 1–43.

Geiser, S., & Santelices, V. (2004). The role of advanced placement and honors courses in college admissions. In P. Gandara, G. Orfield, & C. Horn (Eds.), *Expanding opportunity in higher education: Leveraging promise* (pp. 75–113). Albany: State University of New York.

Geiser, S., & Studley, W. R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment, 8*(1), 1–26.

Grissmer, D. W. (2000). The continuing use and misuse of SAT scores. *Psychology, Public Policy, and Law 6*(1), 223–232.

Hyman, J. (2013). *ACT for all: The effect of mandatory college entrance exams on postsecondary attainment and choice* [Working paper]. Ann Arbor: University of Michigan.

Klasik, D. (2013). The ACT of enrollment: The college enrollment effects of state-required college entrance exam testing. *Educational Researcher, 42*(3), 151–160.

Kling, K. C., Noftle, E. E., & Robins, R. W. (2013). Why do standardized tests underpredict women's academic performance? The role of conscientiousness. *Social Psychological and Personality Science, 4*(5), 600–606.

Kobrin, J. L., Patterson, B. F., Barbuti, S. M., Mattern, K. D., & Shaw, E. J. (2008). *Validity of the SAT for predicting first-year college grade point average* (Research Report No. 2008-5). New York, NY: The College Board.

Lemann, N. (1999). *The big test: The secret history of the American meritocracy*. New York, NY: Farrar, Straus and Giroux.

Morgan, R. (1989). *Analysis of the predictive validity of the SAT and high school grades from 1976 to 1985* (Report No. 89-7). New York, NY: College Entrance Examination Board.

National Center for Education Statistics. (2014). *The Integrated Postsecondary Education Data System: 2012 institutional characteristics* [Data center]. Retrieved from http://nces.ed.gov/ipeds /datacenter/DataFiles.aspx

Noble, J., & Sawyer, R. (2002). *Predicting different levels of academic success in college using high school GPA and ACT composite score* (ACT Research Report Series 2002-4). Iowa City, IA: ACT.

Rothstein, J. M. (2004). College performance predictions and the SAT. *Journal of Econometrics, 121*(1), 297–317.

Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., & Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance? *Psychological Bulletin, 135*(1), 1–22.

Sanchez, E. D. (2013). *Differential effects of using ACT college readiness assessment scores and high school GPA to predict first-year college GPA among racial/ethnic, gender, and income groups* (ACT Research Report Series 2013-4). Iowa City, IA: ACT.

Santelices, M. V., & Wilson, M. (2010). Unfair treatment? The case of Freedle, the SAT, and the standardization approach to differential item functioning. *Harvard Educational Review, 80*(1), 106–134.

Stricker, L. J., Rock, D. A., & Burton, N. W. (1993). Sex differences in predictions of college grades from Scholastic Aptitude Test scores. *Journal of Educational Psychology, 85*(4), 710–718.

Zwick, R. (2012). The role of admissions test scores, socioeconomic status, and high school grades in predicting college achievement. *Pensamiento Educativo, 49*(2), 23–30.

Zwick, R., Brown, T., & Sklar, J. C. (2004). *California and the SAT: A reanalysis of University of California admissions data.* Berkeley: Center for Studies in Higher Education, University of California.

Zwick, R., & Himelfarb, I. (2011). The effect of high school socioeconomic status on the predictive validity of SAT scores and high school grade-point average. *Journal of Educational Measurement, 48*(2), 101–121.